How do proteins find and bind their sites on DNA?

Leonid A. Mirny
Harvard-MIT Division of Health Sciences and Technology, and Department of Physics, Massachusetts Institute of Technology, Cambridge, MA

THE regulation of gene expression relies on transcription factors (TFs) binding to cognate DNA sites. To bind its cognate site, a TF has to recognize it among ~10^6 alternative sites in bacteria or ~10^9 sites in eukaryotes. Using information theory, we ask whether individual TFs possess enough information for such precise recognition. Our analysis of 950 TF motifs gives two strikingly different answers for bacteria and eukaryotes. Bacterial TFs contain sufficient information to recognize cognate sites in their genomes, while eukaryotic TFs are much less specific and do not posses sufficient information. This information deficiency has a profound biological implication: the widespread binding of eukaryotic TFs to thousands of spurious sites on accessible DNA. This promiscuity of eukaryotic TFs can be advantageous for combinatorial regulation, modular use of TFs and, potentially, more evolvable gene regulation.

In another study we examine interactions of TFs with nucleosomes. While nucleosomes limit the access of TFs to DNA, we show that they can also mediate cooperative binding by non-interacting TFs. We find that the mechanism of nucleosome-mediated cooperativity is mathematically identical to the Monod-Wyman-Changeux (MWC) model of cooperativity in hemoglobin. This surprising analogy provides deep insights, with parallels between heterotropic regulation of hemoglobin and roles of nucleosome-positioning sequences and chromatin modifications in gene expression.

[1] Wunderlich Z, and Mirny L.A. Fundamentally different strategies for transcriptional regulation are revealed by information-theoretical analysis of binding motifs http://arxiv.org/abs/0812.3910
[2] Mirny L.A. Nucleosome-mediated cooperativity between transcription factors http://arxiv.org/abs/0901.2905